1 INTRODUCTION

Protein dynamics

The study of protein dynamics by computer simulations

A large diversity of processes in living organisms critically depend on protein activity. Though in many of these processes the mere structure of a protein dominates its function (e.g. collagen in tissues or α-keratin in hair), protein dynamics is crucial to many others. Virtually all biological processes that involve motion find their origin in protein dynamics. Muscle contraction, for instance, is based on the combined action of actin and myosin. Other examples are the molecular motors kinesin and F1-ATPase. Dynamics also plays an important role in many proteins of which the primary function is not mobility itself. For example, the ability to change conformation is also essential for the function of many transport proteins, proteins involved in signal transduction, proteins in the immune system, and numerous enzymes[1]. In many enzymes, conformational changes serve to enclose the substrate, thereby preventing its release from the protein and ideally positioning it for the protein to perform its function, as in lysozyme. Immunoglobulins are highly flexible in order to be able to deal with a large range of ligands. Another role of protein dynamics is found in G-proteins, binding of a hormone to its receptor triggers the dissociation of the α domain from the rest of the protein after a GTP-mediated conformational change. A special class of conformational transitions are found in so-called allosteric proteins. Substrate binding to one subunit of these multimeric proteins triggers a conformational change that alters the substrate affinity of the other subunits, thereby sharpening the switching response of these proteins.

The conformational changes involved range from very subtle, local changes, as in the case of e.g. myoglobin, to global conformational changes, involving motions of significant amplitude for large parts of a protein (e.g. haemoglobin)[1]. Dynamics plays an important role not only in the functional, native state of many proteins, but also the mechanism by which a protein reaches that native conformation, the protein folding process, is a highly dynamic process.

Although a large part of the current knowledge of conformational flexibility in proteins is derived from experimental data (especially X-ray crystallography and Nuclear Magnetic Resonance (NMR)), there is currently no experimental technique that allows monitoring of protein conformational changes at atomic resolution as a function of time at time-scales of nanoseconds. There are several examples of proteins structurally characterised when trapped in different functional states (for an overview, see ref. 2), and the time resolution of structural studies improves steadily[3]. Nevertheless, details on the pathways between different known conformations often remain obscure. Until

now, computer simulation techniques provide the only possibility to obtain dynamic information on proteins at atomic resolution in the picosecond to microsecond time range.

Molecular Dynamics

Out of all possible ways of simulating protein motions, Molecular Dynamics (MD) techniques are among the most popular. In MD, an attempt is made to describe the time evolution of molecular systems as realistically as possible. In a typical simulation, a starting configuration is generated from an experimentally determined structure, and put in an environment that best mimics its natural environment. Obviously, the quality of the obtained dynamic model depends on the quality of the starting model. Once an appropriate starting configuration has been obtained, the actual simulation can be started. In most cases, all particles are treated classically, leaving the problem of solving Newton's equations of motion:

$$\boldsymbol{F}_i = m_i \boldsymbol{a}_i \tag{1.1}$$

with $\boldsymbol{F}_i$ the force, m_i the mass and $\boldsymbol{a}_i$ the acceleration of particle i. Atomic positions $\boldsymbol{x}$ are obtained from:

$$\boldsymbol{a} = \frac{d^2\boldsymbol{x}}{dt^2} \tag{1.2}$$

by numerical integration. At every integration step $\boldsymbol{F}$ is evaluated using:

$$\boldsymbol{F} = -\frac{d\boldsymbol{V}}{d\boldsymbol{x}} \tag{1.3}$$

The potential energy $\boldsymbol{V}$ typically includes terms for covalent bond lengths, angles, torsion angles (dihedrals), improper dihedrals (to maintain tetrahedral or planar geometries), and a number of non-bonded terms[4–6]. The non-bonded terms typically consist of a Lennard-Jones term and an electrostatic (Coulomb) contribution, and in some cases an explicit hydrogen-bonding term. Due to the lack of quantum-mechanical terms, specific parameters must be specified for each atom type in each chemical environment. This results in a parameter-set (force-field) that contains many hundreds of parameters. The absence of polarisability in classical force-fields restricts the reliability of MD simulations, especially in systems where polarisability effects are known to play an important role, as for example in ion-binding proteins. Another potential source of artifacts is the calculation of long-range non-bonded forces.

In relatively large molecular systems (tens of thousands of particles) the combinatorial problem of calculating all pairwise interactions makes the force calculations required for MD simulations extremely time-consuming. The next section gives an overview of techniques proposed to alleviate this problem.

Enhanced efficiency methods

Overview

A clear gap exists between time scales that can currently be obtained by computer simulation techniques applied to biological macromolecules and the times required for most biological processes. With current state of the art methods and computers, a typical protein of 1000 amino-acids (100 kD) can be simulated for time-scales of at most nanoseconds[7], whereas most biological processes take place at times ranging between microseconds to seconds (or even minutes). Even if the rate of increase in computer power (an order of magnitude every 5-7 years[7]) continues, simulation of such processes at the required time scales will be beyond those of standard Molecular Dynamics simulation protocols in the next decade. Therefore, several groups have worked on developing techniques to overcome this problem. Conceptually, three categories of techniques can be distinguished [1]: (i) those that aim to mimic biological systems as realistically as possible and focus on sophisticated (mathematical) methods to enhance computational efficiency, affecting the dynamics as little as possible, (ii) those that simplify the molecular models involved, thus gaining computation time by neglecting details and (iii) those that make use of special properties of the simulated system to describe the system in more appropriate, internal coordinates. This division is not exclusive; some methods cannot be assigned to either category whereas others are hybrid methods based on principles from more than one category. A number of examples from each of the categories will be discussed in this section, and in the next section a technique from the third category, the so-called Essential Dynamics technique, will be described in detail since it will play a key role throughout the rest of this thesis.

Methods to speed up Molecular Dynamics with minimal perturbation

Since the first published application of MD to biomolecular systems[9], a little more than 20 years ago, people have devised methods to increase the time scales of Molecular Dynamics simulations. When Newton's equations of motion are integrated, the limiting factor that determines the time step that can be taken is the highest frequency that occurs in the system. In solvated biological macromolecules, the vibrations of bonds involving hydrogen atoms form the highest frequency vibrations. The bond stretching frequency of an O-H bond is typically about 10^{14} Hz, so the average period would be in the order of 10 fs ($10 \cdot 10^{-15}$ s)[10]. This limits the time-step to be taken in MD simulations to about 0.5 fs (a rule of thumb exists that states that for a reasonable sampling of a periodic function, samples should be taken at least twenty times per period). The introduction of a method to constrain these bonds (or, in fact, all covalent bonds) allowed to increase the time step to a

[1] Previously, a subdivision has been suggested according to levels of approximation[8]

typical value of 2 fs[11]. Since these bond vibrations are practically uncoupled from all other vibrations in the system, constraining them does not notably alter the rest of the dynamics of the system. This is not true, however, for bond-angle fluctuations, which form the second-highest frequency vibrations. Constraining bond-angles has a severe effect on many other fluctuations in the system, including even global, collective fluctuations, limiting the use of methods that use bond-angle constraints to only a few specific cases[10].

The notion that a number of discrete classes of frequencies of fluctuations in simulations of biomolecules can be distinguished, however, can be utilised to design more efficient algorithms. Forces that fluctuate rapidly need to be recalculated at a higher frequency than those that fluctuate on a much longer time scale. Although not trivial to implement, a number of successful applications of so-called multiple time-step algorithms have been reported in the literature (for a review, see ref. 10). Speedup factors of 4-5 have been claimed for such methods with respect to unconstrained dynamics, making them only slightly more efficient than simulations with covalent bond-length constraints.

As stated before, the most time-consuming part of Molecular Dynamics simulations is the force evaluation at every time-step. Especially the evaluation of electrostatic forces is notorious since Coulomb terms are inversely proportional to the inter-atomic distances of charged particles. This makes their contribution to the total force non-negligible even at fairly large distances (above 10 Å). Several methods have been proposed to reduce the computational cost to calculate long-range electrostatic forces. The most straightforward of these methods are cut-off methods where interactions beyond a certain radius are simply neglected[12]. This reduces the original order of complexity from N^2 to N (with N the number of particles) but significant artifacts have been reported at the edge of the cut-off radius[13–15]. Ewald methods form the traditional way to calculate electrostatic interactions in a more elegant fashion by calculating infinite lattice sums, but the order of complexity $N^{3/2}$ makes the method unsuitable for simulation of large biomolecular systems. However, approximations like particle-particle particle-mesh (PPPM)[16] and particle-mesh Ewald (PME)[17] that scale with $N \cdot \log(N)$, have shown encouraging results[18,19]. Fast Multipole methods (FMM) distribute atomic charges over a hierarchy of clusters and approximate electrostatic interactions by multipolar expansions of the potential generated by the clusters[20]. FMM methods even scale with N but require extra overhead compared with other methods, making them the method of choice only for systems of tens or hundreds of thousands of particles. Combinations of efficient ways to calculate electrostatic interactions with multiple time-step methods have already been described (e.g. FMM together with multiple time step algorithms[21,22]).

Another approach to reach equilibrium conformational properties at an enhanced rate is by performing so called 'mass tensor Molecular Dynamics'[23]. The masses of e.g. hydrogen atoms are increased to slow down the highest-frequency vibrations, allowing for a larger integration step. The dynamics

is perturbed in this way, but equilibrium properties are not affected[24]. Another way to get around the problem of high frequency vibrations of hydrogen atoms is by excluding them from the actual integration and regenerating their positions every time step from the positions of the heavy atoms to which they are attached[25]. Although the features of this approach have yet to be explored, initial results have shown that a time-step of 6 to 8 fs is within reach. Another approach has recently been proposed, called "self-guided Molecular Dynamics"[26], that introduces an additional systematic force that is based on earlier parts of the simulation. Enhanced rates of conformational sampling have been claimed for small peptides. Its applicability in the field of protein dynamics still needs to be studied.

Simplified protein models

Before the first all-atom Molecular Dynamics simulation on a protein was performed, simulations of protein folding with a simplified protein model had been reported[27]. This illustrates the limitations of all-atom descriptions of proteins in computer simulations, especially in the presence of explicit solvent.

Simplified protein models have been utilised extensively in the field of protein folding. Employed methodologies include lattice Monte Carlo (MC) models and adapted MD or Langevin Dynamics (LD) models. The Monte Carlo technique is a stochastic method: random displacements are taken at each step, which are only accepted when an energy criterion is fulfilled[28]. Lattice models form perhaps the most simplified models with some resemblance to real proteins[29]. Their advantage is that exhaustive searches of the configuration space can be reached for small proteins (up to about 100 residues) by MC methods[30–33]. However, their applicability is limited due to the lack of detail in the models and the restriction of the search space due to lattice constraints. Continuum models of simplified proteins (bead models) utilising adapted MD or LD algorithms are more promising, in that sense, because of the absence of lattice restrictions. In Langevin Dynamics, compared to eq. 1.1, forces contain an additional friction and noise term to mimic the effect of solvent (which is not treated explicitly)[34]. Although exhaustive searches can usually not be reached by these bead methods, promising results have been reported[8,35–37]. Another application of simplified protein models for use in protein folding are so called threading techniques[38] (for recent reviews, see refs. 39, 40). The idea is that a discrete number of folds exists to which proteins are restricted. The sequence of a protein with unknown structure is threaded through a set of known protein folds, after which suitable scoring potential (e.g. ref. 41, for a review see ref. 42) reveals which structure is most probable for that sequence.

Monte Carlo calculations using coarse grained protein models similar to those used for threading have shown that native state dynamics of proteins can successfully be simulated at a rate one order of magnitude faster than can be obtained by all-atom models[43,44]. Also, LD simulations with a multiple time step algorithm showed vast improvements of computational efficiency

compared to traditional MD, even when using an all-atom representation[45]. Apart from the advantage of a multiple time-step algorithm, part of the computational efficiency in this model is the result of the absence of explicit solvent molecules. Several methods of solvent treatment by implicit models have been suggested over the years[46–50], but their range of applicability is still a matter of debate[51–54]

Simplification in its most extreme form reduces a protein's conformational space to that of two or more rigid bodies. Domain motions are known to form the basis of the function of several proteins (see e.g. ref. 2) and therefore many properties of the functional mechanism of such proteins may be studied by focusing on the rigid-body motions of the domains involved[55,56]. Even in single-domain proteins, quasi-rigid parts have been identified (for example secondary structure elements[57–59]). This observation could in principle be used in a simplified protein model, but has so far only been applied in the field of theoretical protein folding[60,61].

Protein dynamics in internal coordinates

Efficiency of computer simulations can be enhanced by describing the simulated systems in their internal degrees of freedom, as opposed to the usual Cartesian coordinates. The goal, as in the previous section, is to reduce the number of degrees of freedom in the simulated system. The methods described in this section, however, retain the atomic detail of the modeled system. Perhaps the first example of this method was proposed by Ryckaert & Bellemans[62] in their simulation of n-butane, with only one internal degree of freedom (the central torsion angle). For proteins, the use of torsion angles also seems an appropriate choice since dihedral angles are the main degrees of freedom, of which the ϕ and ψ backbone dihedrals play the largest role in large-scale protein motions. Application of torsion angles in the study of protein dynamics has been proposed for MC[63] and MD[64] simulations. The advantage of such techniques is that larger simulation steps (either time-steps in MD or space-steps in MC) can be taken in the simulation. Stable MD simulations with time steps of 13 fs have been described for an Ala_9 peptide[64], whereas time-steps of at most 2 fs can be taken when only bond lengths are constrained. However, a number of problems is encountered when protein dynamics is described in torsion angle space. First, when the equations of motion are solved for these internal coordinates, the inverse of the moments of inertia tensor is required every time step. Since matrix inversion scales with the third power of the number of matrix elements in terms of computation time, application of such methods is limited to small systems. However, a method to get around this problem has been proposed[65], reducing the computational cost to order N instead of N^3. The second problem connected with torsion-angle dynamics is the absence of bond-angle fluctuations. Bond-length fluctuations can safely be neglected, but constraining bond-angles severely restricts dynamics of proteins (see e.g. ref. 10). Due to the altered potential employed in torsion-angle approaches, conformational barriers are overestimated, making

the method most useful for simulations at elevated temperatures, used for example in the field of refinement of NMR structures[66].

Torsion angle approaches have also been applied in combination with knowledge-based force-fields. Monte Carlo simulations have been reported claiming enhanced convergence for NMR structure determination[67]. Also in off-lattice simulations MC calculations in torsion-angle space have begun to gain popularity (for a review, see e.g. ref. 68).

Another way to define internal coordinates in proteins is based on the notion that most positional fluctuation occurs along collective degrees of freedom. This was first realised from Normal Mode analyses of a small protein[69–71]. In Normal Mode analyses, the potential energy surface is assumed to be harmonic. Collective variables are obtained by diagonalisation of the Hessian matrix (second derivative of the potential energy) in a local energy minimum. Quasi harmonic analysis[72–75], principal component analysis[76–78] and singular value decomposition[44,79] of Molecular Dynamics trajectories of proteins have shown that even beyond the harmonic approximation, protein dynamics is dominated by a limited number of collective coordinates. These methods seek those collective degrees of freedom that best approximate the total amount of fluctuation. The subset of largest-amplitude variables form a set of generalised internal coordinates that can be used to effectively describe the dynamics of a protein. As opposed to torsion angles as internal coordinates, these collective internal coordinates are not known beforehand. Unless many experimental structures are available, a simulation is required to obtain a definition of these coordinates. Once an approximation of the collective degrees of freedom has been obtained, simulations in the space spanned by only these coordinates can in principle be initiated. Such a technique has successfully been applied to small molecules[80]. However, coupling of the main modes of collective fluctuation to more constrained coordinates is likely to be responsible for a limited applicability in dynamic simulation of proteins (e.g. ref. 10 and A. Amadei and T. Linssen, personal communication). Methods to bypass the problems of this coupling include biased MD simulations with constraints along collective internal coordinates derived from earlier simulations[81] and form the subject of chapters 3 and 4 of this thesis. The dynamics can also be biased by modifying the potential energy function along such a collective degree of freedom. This is thought to be especially useful for enhancing the rate of conformational transitions in proteins[82].

Essential Dynamics

The Essential Dynamics (ED) technique is a method from the third category of the last section. A brief description will be given here, discussing some important features of the method. For a more rigorous description, see ref. 78. As an analysis technique, ED is based on a principal component analysis of (MD generated) structures. A principal component analysis is a multidimensional linear least squares fit procedure. To understand how this is

applicable to protein dynamics, the usual three-dimensional (3D) Cartesian space to represent protein coordinates (which is e.g. used to represent protein conformations in the Brookhaven Protein Data Bank or PDB) needs to be replaced by another, multidimensional space. A molecule of N particles can be represented by N points in 3D space. With 3 coordinates per point, this adds up to 3N coordinates. In a 3N-dimensional space, however, such a structure can be represented by a single point. In this space, this point is characterised by 3N coordinates. This representation is convenient since a collection or trajectory of structures can now be regarded as a cloud of points. Like in the case of a two-dimensional cloud of points, also in more dimensions, always one line exists that best fits all points. As illustrated for a two-dimensional example (Fig. 1.1), if such a line fits the data well, the data can be approximated by only the position along that line, neglecting the position in the other direction. If this line is chosen as coordinate axis, then the position of a point can be represented by a single coordinate. In more dimensions the procedure works similarly, with the only difference that one is not just interested in the line that fits the data best, but also in the line that fits the data second-best, third best, and so on (the principal components). These directions together span a plane, or space, and the subspace responsible for the majority of the fluctuations has been referred to as the 'essential subspace'. Applications of such a multidimensional fit procedure on protein configurations from MD simulations of several proteins has proven that typically the ten to twenty principal components are responsible for 90 % of the fluctuations of a protein[76–78]. These principal components correspond to collective coordinates, containing contributions from every atom of the (protein) molecule. Summarised, a limited number of collective motions is responsible for a large percentage of a protein's conformational fluctuations.

If all atoms in a protein were able to move uncorrelated from each other, an approximation of the total fluctuation by only a few collective coordinates would not be possible. The fact that such an approximation is successful is the result of the presence of a large number of internal constraints and restrictions ('near-constraints') defined by the interactions present in a given protein structure. Atomic interactions, ranging from covalent bonds (the tightest interactions) to weak non-bonded interactions, together with the dense packing of atoms in native-state protein structures form the basis of these restrictions.

In the study of protein dynamics, only internal fluctuations are usually of interest. Therefore, the first step in an Essential Dynamics analysis is to remove overall rotation and translation. This is done by translation of the center of mass of every configuration to the origin after which a least squares rotational fit of the atoms is performed onto to a reference structure. Recently it was suggested that this procedure might lead to a bias in the definition of the internal fluctuations, and that a way to circumvent this bias would be to work in distance space[83]. The actual principal component analysis is based on construction and diagonalisation of the covariance matrix of positional fluctuations. The covariance matrix is constructed from the

atomic coordinates according to:

$$\boldsymbol{C}_{ij} = \langle (\boldsymbol{x}_i - \langle \boldsymbol{x}_i \rangle)(\boldsymbol{x}_j - \langle \boldsymbol{x}_j \rangle) \rangle \tag{1.4}$$

where $\boldsymbol{x}$ represents the atomic coordinates and the angle brackets a time or ensemble average. Particles moving in a correlated fashion correspond to positive matrix elements (positive correlation) or negative elements (negative correlation), and those that move independently to small matrix elements. The orthogonal transformation $\boldsymbol{T}$ that diagonalises this (symmetric) matrix contains the eigenvectors or principal components of $\boldsymbol{C}$ as columns and the resulting diagonal matrix $\boldsymbol{\Lambda}$ contains the corresponding eigenvalues:

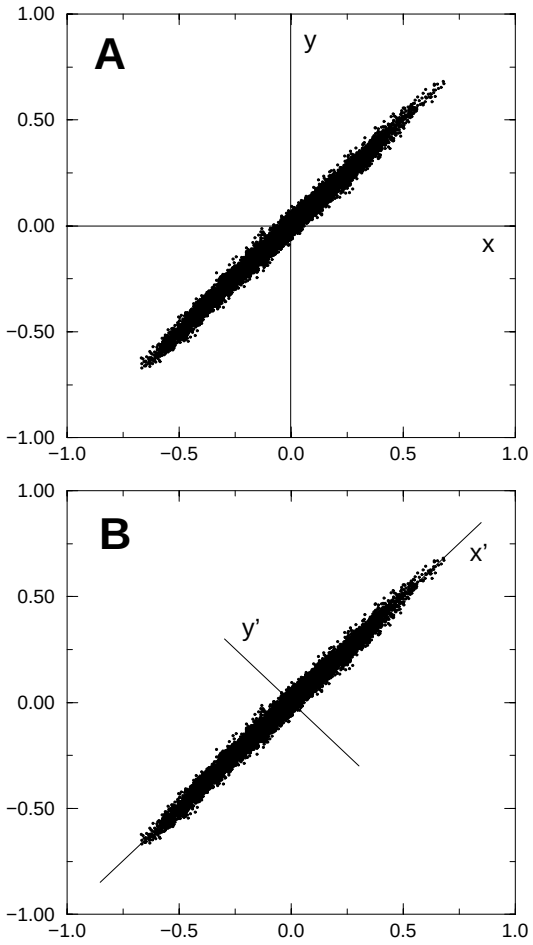

Figure 1.1 Illustration of Essential Dynamics in two dimensions. With a distribution of points as depicted here, two coordinates (x,y) are required to identify a point in the cluster in panel A, whereas one coordinate (x') approximately identifies a point in panel B.

$$\mathbf{\Lambda} = \boldsymbol{T}^{\top}\boldsymbol{C}\boldsymbol{T} \tag{1.5}$$

The eigenvalues are a measure of the mean square positional fluctuation along the corresponding eigenvector. When the eigenvectors are sorted to decreasing eigenvalue, the first eigenvectors are those collective motions that best approximate the sum of fluctuations and the last eigenvectors correspond to the most constrained degrees of freedom. The characteristics of these collective fluctuations can be studied by projecting the ensemble of structures onto single eigenvectors and by translation of these projections to 3D space to visualise the atomic displacements connected with that eigenvector. As stated above, analyses of MD trajectories of several proteins have shown that few collective coordinates dominate the dynamics of native proteins (together often referred to as the 'essential subspace'). In a number of cases these main modes of collective fluctuation were shown to be involved in the functional dynamics of the studied proteins[78,84–86].

ED analyses can be applied to any subset of atoms of the ensemble of structures[78] and are not restricted to ensembles generated by MD simulation. Applications to collections of X-ray structures[86,87], NMR structures[88] and structures derived from distance constraints[89] have been reported. Since collective (backbone) fluctuations dominate the dynamics of proteins, usually only backbone or C-α coordinates are used to save computation time and to prevent problems with apparent correlation of side chain motions with backbone motions which are merely the result of poor statistics. However, even when the method is applied to only C-α atoms, the diagonalisation of the covariance matrix can still be an enormous computational task. An approximation has been developed to alleviate this problem, allowing analyses of systems with thousands of amino-acids[90].

Although first designed for proteins, the ED method can in principle be applied to any constrained (biomolecular) system. Successful applications to DNA have already been reported[91,92].

Identification of the dominant modes of collective fluctuation is the first step in the Essential Dynamics technique. As sketched in the previous section, knowledge of the essential subspace can be used in a sampling technique that exploits the limited dimensionality of that space to achieve a more efficient sampling than can be obtained by more conventional techniques.

Outline of this thesis

The second chapter of this thesis is concerned with the convergence of Essential Dynamics results from relatively short MD simulations. In the literature, it had been reported that principal component analysis (Essential Dynamics is a principal components analysis of the atomic fluctuations) of MD simulations of such short time lengths is not suitable for describing long-time scale

protein dynamics because this subspace keeps changing throughout the simulations. Apart from the issue of convergence of the essential subspace, the sensitivity of the essential dynamics results to MD parameters is also examined in this chapter. A set of reference simulations is compared to a set in which parameters were modified that were believed to have a potential effect on (protein) dynamical or configurational properties.

The third chapter presents an extension of the Essential Dynamics sampling technique. This ED sampling technique is based on the idea that, since most fluctuations in proteins take place in a hyperspace of limited dimension, a systematic or otherwise enhanced sampling in this subspace will result in an efficient way to explore the configurational space of proteins. A prerequisite for success of this method, of course, is a sufficiently accurate approximation of the subspace. In a first implementation, the method had yielded encouraging results on a small protein which showed that indeed acceptable protein structures were generated which were more widely spread in configuration space than would be obtained by usual MD simulation[81]. This chapter presents the application of a modified sampling algorithm to a peptide hormone. An extensive sampling is performed and the stability of resulting structures is measured by subjecting them to MD simulation without essential dynamics constraints. Based on these results, a model is presented for the free energy surface of this peptide and proteins in general in the space of the major collective conformational coordinates.

Encouraged by the results on the peptide, the ED sampling procedure was applied to a small protein: the Histidine containing Phosphocarrier protein HPr. It was found that some modifications to the algorithm were required because denaturation of the protein was observed when the same criteria were used as with the peptide. In chapter 4, the resulting ensemble of structures is compared to a set of structures collected from unconstrained MD simulations and from simulations with NMR-NOE restraints. Structures extracted from the latter simulations represent the high-resolution NMR structure of HPr[93]. NOE violations from each of the three runs are compared to each other, as well as several geometrical and energetical properties.

Chapter 5 presents a comparison of domain motions in T4 lysozyme calculated from several crystal structures and those obtained from MD simulation. T4 lysozyme is among the best experimentally characterised proteins in terms of conformational properties and therefore is an ideal candidate for a rigorous test how well MD/ED results from simulations in the order of nanoseconds correspond to known, large-scale collective fluctuations in proteins. A newly developed method to characterise domain motions in proteins was employed to compare the experimental and theoretical results in detail. Not only methodological implications, but also functional aspects of the domain fluctuations are described.

The observation that most of a protein's positional fluctuation can be approximated by only a few collective degrees of freedom led to an attempt to derive those degrees of freedom by another, computationally less demanding

method. The restriction of a protein's fluctuations to a hyperspace of limited dimension is caused by the presence of a large number of explicit and implicit constraints and restrictions to the configurational freedom of each atom. The idea arose that if the network of interactions responsible for these restrictions could be represented in a simpler way than in e.g. MD simulations, an approximation of the constraint surface, and therefore also of the complementary essential subspace could be obtained. Chapter 6 introduces a technique, named CONCOORD, that generates protein structures within predefined distance bounds. CONCOORD structures of different proteins are compared to structures generated by MD, in terms of Essential Dynamics properties and more conventional techniques.

Chapter 7 presents an application of the CONCOORD method to the molecular chaperonin GroEL. The elucidation of the X-ray structures of GroEL in different conformations together with electron microscopy data had shown that GroEL is a remarkably flexible protein and that allosteric properties play an important role in its function: to assist other proteins to fold to their native conformation. The size of the protein (M≈800kD) makes it unsuitable for other computational techniques that yield protein conformational properties, such as MD, but because of its algorithmic simplicity and efficiency, it proved possible to apply CONCOORD. Essential dynamics analyses were applied to the collection of experimental structures and conformations generated by CONCOORD. Previously unnoticed features of the crystallographic structures are presented, in combination with conformational properties derived from the CONCOORD simulations. Implications for the allosteric mechanism of GroEL are described.

Finally, chapter 8 finishes this thesis with some concluding remarks on theoretical approaches in the field of protein dynamics and an outlook to the future.